Phonotactic Probability

Phonotactic Probability Effects in Polysyllabic Nonwords

A common means of assessing immediate processing of a novel phonological word form is to ask individuals to repeat the form, as in the nonword repetition task. There has been increasing interest in the nonword repetition task over the past two decades, originating with demonstrations of a relationship between an individual’s ability to repeat nonwords and the ability to learn new words (Gathercole & Baddeley, 1989). The processes involved in nonword repetition share much in common with novel word form learning. As described by Gathercole (2006), both nonword repetition and novel word form learning must undergo auditory processing, phonological analysis, and phonological storage. Nonword repetition can potentially reveal a great deal about the initial processing and eventual learning of novel word forms.

Initially the nonword repetition task was thought to be a relatively “pure” measure of phonological short-term memory (Gathercole & Baddeley, 1989), although there have since been numerous demonstrations of the effects of prior knowledge on nonword repetition abilities (Edwards, Beckman, & Munson, 2004). Given that prior knowledge seems to have an influence on the ability to repeat nonwords, it becomes important to understand exactly how this knowledge affects processing. The most common way in which similarity of a nonword to known words has been manipulated is using a measure known as phonotactic probability (Jusczyk, Luce, & Charles-Luce, 1994).

Phonotactic probability is most commonly calculated as the frequency with which a phoneme, or sequence of phonemes, occurs within existing words of a language (Vitevitch & Luce, 2004). This type of manipulation of similarity of a nonword has been used to investigate nonword repetition accuracy (Munson, Edwards, & Beckman, 2005), lexical decision response times (Vitevitch & Luce, 2005), and novel word learning (Storkel, 2001). Understanding how
Phonotactic Probability

individuals are assessing similarity in these situations can help reveal information about the nature of the representations used in these processes. Gupta & Tisdale (2009) offered a computational account of many of these effects, showing how these phenomena fall out of an architecture in which serial order processing and linguistic knowledge are closely integrated. Nevertheless, a number of questions remain regarding how linguistic knowledge affects nonword repetition.

In particular, phonotactic probability is often manipulated for whole word forms, although the calculation of word form phonotactic probability relies on a summation of values from individual phones and biphones. While this manipulation has been used repeatedly to demonstrate that high phonotactic probability is beneficial for repetition accuracy (Munson, 2001; Vitevitch, 2003), the manipulations used are typically short one- or two-syllable nonwords. While the findings have been fairly consistent that high probability sequences are beneficial for repetition it is not clear whether the benefits of these probability manipulations arise at a local (within-syllable) level or can have more global effects on other syllables in the nonword which are not manipulated. Answering this question about the location of these similarity effects may have implications for the type of representation involved in the processing and repetition of nonwords.

Because most previous work in phonotactic probability manipulations has been conducted by manipulating an entire word form’s similarity to known words it is difficult to assess what exactly is producing these beneficial effects. For example, it could be that specific serial positions within a nonword produce different effects of similarity that are masked by the previously discussed whole-word manipulations. A more systematic manipulation and measurement of the effects of similarity at various points within a nonword could allow for these
Phonotactic Probability

distinctions to be made. It may be the case that similarity has different effects across different positions within a nonword.

In an attempt to begin to address these questions about the nature of similarity effects in nonword repetition, multiple experiments were conducted to examine the effects of manipulating a single syllable within a longer nonword. All of the nonwords used in the present experiments were constructed with a five-syllable structure. Using such a structure it was possible to create nonword “frames” to serve as the context for a single-syllable manipulation. Each frame had both a high and a low probability syllable which could be attached to it to complete the five-syllable structure. The question driving this study was whether or not manipulating a single syllable within a longer nonword would have primarily local effects or if similarity at any given point could have more widespread effects on repetition accuracy. In addition, the structure of the nonwords allowed us to ask if these local, or potentially global, effects would differ in some substantial way in different syllable positions within nonwords.

**Experiment 1: First Syllable Manipulation**

In this experiment, nonwords were created that varied between conditions only in the first syllable. The effects of manipulating the first syllable of otherwise identical nonwords might be expected to have primarily local effects given that only the first syllable differed between nonwords. It is also possible that even this local manipulation might have an effect on repetition performance on syllables shared between nonword frames. We predicted that high similarity of the manipulated syllables would result in higher repetition accuracy for those syllables but did not have strong predictions for the effects on any following syllables.

**Method**
Phonotactic Probability

**Participants.** 16 undergraduate psychology majors at the University of Iowa participated in this experiment in exchange for partial fulfillment of course requirements. All participants were native English speakers and had normal hearing and either normal or corrected-to-normal vision.

**Materials and design.** Nonwords used in the present experiment were specifically created to vary phonotactic probability of the first syllable while the remaining four syllables were identical. Each nonword was composed of four consonant-vowel syllables (CV) with a final consonant-vowel-consonant (CVC) syllable. An online calculator for computing phonotactic probabilities was used to determine the values for all possible CV and CVC combinations in each of the five syllable positions (Vitevitch & Luce, 2004). The second through fifth syllables for each nonword were chosen such that two possible first syllables could be paired with it, creating two separate nonwords with identical second through final syllables. These identical sections of two paired nonwords will be referred to as “frames.” This resulted in 64 nonwords (composed of 32 individual frames), 32 of which had initial syllables high in phonotactic probability and 32 of which had initial syllables low in phonotactic probability. For example, the nonwords “vobimapariv” and “tybimapariv” have different first syllables but share a common second- through fifth-syllable frame. These nonwords were then recorded by a male native English speaker at a sampling rate of 44.1kHz. All nonwords had primary stress on the fourth syllable with secondary stress on the second syllable.

Each subject was asked to repeat a given frame only once, resulting in lists of 32 nonwords to be repeated by each subject. Half of the frames on each list were paired with a high probability first syllable and half were paired with a low probability first syllable. These pairings were counterbalanced across participants. Comparisons within participants can then be made.
Phonotactic Probability

regarding differences in performance on high- versus low-probability first syllables and comparisons across participants can be used to examine the effects of high- versus low-probability first syllables on the shared nonword frames.

**Procedure.** Participants listened to nonwords through headphones in a sound treated room. Stimulus presentation was controlled using PsychoPy version 1.76.00 (Peirce, 2007) and all nonwords were presented auditorily. Participants were asked to repeat each nonword after hearing the sound were given four practice trials using stimuli not present in the rest of the experiment. During each trial a fixation cross was presented in the center of the screen for 500ms. At the offset of the fixation cross a nonword was presented for repetition. A trained research assistant present in the room during the experiment manually began the next trial after an attempt at repetition by the participant. After the research assistant manually indicated that the repetition attempt was complete these was a 1s interval before the next trial began. Trials were arranged randomly into blocks of 16 items and an opportunity to rest was given in between the two blocks. Participant responses were digitally recorded to be transcribed and scored offline. To be considered correct, a given syllable had to be produced with the correct phonemes in the correct serial positions. Trained research assistants transcribed the recordings and were unaware of which nonwords were high or low in phonotactic probability.

**Results and Discussion**

Mean proportion of correct responses by syllable position are available in Figure 1, while mean proportion of correct whole word form repetitions are available in Figure 2.

To analyze the results, a 5 (syllable position) x 2 (phonotactic probability) repeated measures ANOVA was conducted on proportion of correct repetitions using an arcsine transformation. This revealed both a main effect of syllable position, $F(4,60) = 27.60, p < .001,$
Phonotactic Probability

partial $\eta^2 = .648$, as well as phonotactic probability, $F(1,15) = 5.09, p = .039$, partial $\eta^2 = .254$. However there was also a significant interaction between syllable position and phonotactic probability, $F(4,60) = 3.79, p = .008$, partial $\eta^2 = .202$. Analysis of the simple main effect of phonotactic probability at each syllable position revealed that the effect of phonotactic probability was only significant for the first syllable, $F(1,15) = 15.95, p = .001$, partial $\eta^2 = .515$, with syllables two through five showing no significant differences between high and low phonotactic probability conditions, all $p > .10$, all partial $\eta^2 < .166$. Corrections for multiple comparisons were made using the Dunn-Sidak method ($\alpha = .05$).

Because we were also interested in the potential global effects, and because previous studies have generally reported whole word form repetition accuracy in phonotactic probability experiments, we also conducted an analysis of whole word form repetition accuracy. A paired-samples t-test was conducted on the proportion of nonwords successfully repeated based on phonotactic probability condition. Whole words were only scored as correct if all phonemes were correctly produced during the repetition attempt. If only examining whole word repetition accuracy (rather than examining at a more fine-grained level) we would have concluded that high phonotactic probability leads to greater whole word form repetition accuracy, $t(15) = 2.26, p = .04$. While this result is significant it is worth pointing out that this effect (when analyzed at the syllable level) is clearly being driven by the difference in repetition accuracy at the first syllable.

The results of this experiment indicate that manipulating only the first syllable of a nonword, while the remaining four syllables remain constant counterbalanced across participants, produces primarily local effects in nonword repetition accuracy. This suggests that the processing of a nonword may be due to sequencing mechanisms similar to those used in serial list recall. In addition, these results suggest that similarity of nonwords may produce
Phonotactic Probability

effects during this sequential processing thus driving the local effects observed here. It is worth noting, however, that while no other syllable displayed statistically significant differences in repetition accuracy the direction of the effect at each syllable was such that the high probability first syllable manipulation did lead to higher accuracy scores in each case. It could be that a lack of power to detect an effect at these other syllable positions was the cause of these null results.

This experiment manipulated only the first syllable in each presented nonword. While the results do suggest a primarily local effect of the manipulation, it remains to be seen whether or not this effect will hold across syllable positions. To begin to address this question we conducted Experiment 2, which manipulated similarly constructed nonwords only in the second serial position.

**Experiment 2: Second Syllable Manipulation**

This experiment was conducted to examine the effects of local similarity within the second, rather than initial, syllables of nonwords. It may be the case that initial syllable similarity has different effects on nonword repetition than it does for intermediate syllables within a nonword. To address this possibility we manipulated only the second syllables of nonwords. Once again we predicted that second syllables with high phonotactic probability would show an increase in repetition accuracy over those low in phonotactic probability. Just as in Experiment 1 we also made no strong predictions about effects at other syllables.

**Method**

**Participants.** 16 undergraduate psychology majors at the University of Iowa participated in this experiment in exchange for partial fulfillment of course requirements. All participants were native English speakers and had normal hearing and either normal or corrected-to-normal vision.
Phonotactic Probability

**Materials and Design.** Nonwords in Experiment 2 were generated using the same methods as in Experiment 1 with the following differences. Rather than creating nonword frames that contained identical second through final syllables we created nonword frames that were matched on initial syllables as well as syllables three through five. Each frame was paired with two second syllable options, one high in phonotactic probability and the other low in phonotactic probability. Because phonotactic probability varies based on position within a nonword (Vitevitch & Luce, 2004), the possible frames and high and low probability manipulated syllables differed from Experiment 1. As in Experiment 1, there were a total of 64 nonwords created for Experiment 2 (32 high probability and 32 low probability). Paired lists were created just as in Experiment 1.

**Procedure.** The procedure in the present experiment was identical to Experiment 1.

**Results and Discussion**

Mean proportion of correct responses by syllable position are available in Figure 3, while mean proportion of correct whole word form repetitions are available in Figure 4.

To analyze the results, a 5 (syllable position) x 2 (phonotactic probability) repeated measures ANOVA was conducted on proportion of correct repetitions using an arcsine transformation. This revealed both a main effect of syllable position, $F(4,60) = 32.44, \ p < .001$, partial $\eta^2 = .684$, and of phonotactic probability, $F(1,15) = 9.60, \ p = .007$, partial $\eta^2 = .390$. However, unlike Experiment 1 there was no indication of an interaction between syllable position and phonotactic probability, $F < 1$. Given the previous results, however, planned comparisons of the effect of phonotactic probability at each syllable position were conducted, and corrections for multiple comparisons were made using the Dunn-Sidak correction ($\alpha = .05$). These comparisons revealed that the only place phonotactic probability had a significant effect
Phonotactic Probability

on repetition accuracy was at syllable two, $F(1,15) = 6.20, p = .025$, partial $\eta^2 = .292$. All other syllables demonstrated no significant difference between high and low phonotact­ic probability conditions, all $p > .2$, all partial $\eta^2 < .104$.

As in Experiment 1, we also conducted a whole word form level analysis using a paired-samples t-test on the proportion of nonwords successfully repeated by condition. Just as was observed for the first syllable manipulation, if examined only at the whole word level high phonotactic probability seems to lead to more accurate repetition of the entire word form compared to low phonotactic probability, $t(15) = 2.36, p = .032$.

These data still seem to suggest a primarily local effect of phonotactic probability when planned comparisons are taking into account. However, the lack of a significant interaction while both main effects did reach significance is cause enough to consider that local manipulations of phonotactic probability may be having more distal effects than were found in Experiment 1. The effect size of the phonotactic probability manipulation at the manipulated syllable was also greatly reduced in the present experiment compared to the effect observed in Experiment 1. Still, it is worth noting that the effect size of the manipulation within these results is still far larger at the manipulated syllable than at any other again suggesting that even if there are more distal effects the majority of the main effect detected arises from the manipulated syllable and not from other syllable positions.

**General Discussion**

The results of these two experiments form a pattern that in a general sense replicates past findings on the effects of phonotactic probability. In both experiments it was shown that nonwords high in phonotactic probability showed an increase in accuracy of repetition compared to nonwords low in phonotactic probability. What is novel about these findings is that we were
Phonotactic Probability

able to demonstrate a primarily local effect of these manipulations. Given the complexity of five syllable novel word forms for both perceptual processing and speech planning it is quite telling that effects arise at the manipulated syllable that are not detectable at other syllable positions.

Still, it is worth noting that while the differences at non-manipulated syllables were not different statistically, the numerical differences in means were all in the predicted direction had we been expecting a more global effect of phonotactic probability. In addition, if the effect of phonotactic probability were indeed completely local and had no impact on the cohesiveness of the word form as a unit we would not have expected to detect a whole word level effect in both experiments as we did with the present results.

The current findings lead us to conclude that while the direct effects of phonotactic probability manipulations arise at the syllable level, there may be some subtle impact on processing of the entire word form that helps the syllable sequence as a whole. Past research has been unable to examine such effects because of the nature of manipulating entire word forms in phonotactic probability rather than smaller segments, or failure to analyze results at the syllable level. These results suggest an important role of syllable-level representations in nonword processing and word learning.
Phonotactic Probability

Figures

Figure 1. **Proportion of Correct Repetitions by Syllable Position.** The high and low conditions here represent a high or low probability first syllable for each nonword. Nonwords did not differ on any syllable but the first between subjects. Error bars represent one standard error of the mean. * $p = .001$, all others n.s.

Figure 2. **Whole Word Repetition Accuracy for First Syllable Manipulation.** Error bars represent one standard error of the mean. * $p < .05$. 

- **Correct Repetitions by Syllable Position**
- **Whole Word Repetition Accuracy**
Phonotactic Probability

Figure 3. *Proportion of Correct Repetitions by Syllable Position*. The high and low conditions here represent a high or low probability second syllable for each nonword. Nonwords did not differ on any syllable but the second between subjects. Error bars represent one standard error of the mean. *p < .05, all others n.s.

Figure 4. *Whole Word Repetition Accuracy for Second Syllable Manipulation*. Error bars represent one standard error of the mean. *p < .05.*
Phonotactic Probability

References


